

Claims

We claim:

- 1 1. A method for identifying talking heads in a compressed video, comprising:
2 extracting motion activity descriptors from each of a plurality of shots;
3 combining the plurality of motion activity descriptors of each shot, into a
4 shot motion activity descriptor;
5 measuring a distance between the shot motion activity descriptor and a
6 template motion activity descriptor; and
7 identifying a particular shot as a talking head if the measured distance is less
8 than a predetermined threshold.
- 9 2. The method of claim 1 further comprising:
10 extracting a plurality of training motion activity descriptors from a training
11 video including a plurality of training shots, each training shot including a training
12 talking head; and
13 combining the plurality of training motion activity descriptors into the
14 template motion activity descriptor.
- 15 3. The method of claim 2 wherein the combining is a median of the plurality of
16 training motion activity descriptors.

- 1 4. The method of claim 2 wherein the combining is a mean of the plurality of
2 training motion activity descriptors.
- 1 5. The method of claim 1 further comprising:
2 normalizing the measured distance.
- 1 6. The method of claim 1 wherein the threshold is a standard deviation σ of the
2 temple motion activity descriptor.
- 1 7. The method of claim 1 wherein each motion activity descriptor is of the form
2 $C_{mv}^{avg}, N_{sr}, N_{mr}, N_{lr}, \sigma_{fr}$, where C_{mv}^{avg} is an average motion vector, and N_{sr}, N_{mr}, N_{lr}
3 are short, medium and long run zero-length motion vectors, respectively.
- 1 8. The method of claim 7 wherein the distance is measured according to:
2
3
$$D(S, T) = \frac{W_{tot}}{C_{avg}(T)} |C_{avg}(T) - C_{avg}(S)| + \frac{W_{tot}}{N_{sr}(T)} |N_{sr}(T) - N_{sr}(S)|$$

$$+ \frac{W_{tot}}{N_{mr}(T)} |N_{mr}(T) - N_{mr}(S)| + \frac{W_{tot}}{N_{lr}(T)} |N_{lr}(T) - N_{lr}(S)|$$
where W_{tot} is a normalizing weight, T is the temple motion activity descriptor,
3 and S is the shot motion activity descriptor.
- 1 9. The method of claim 1 further comprising:
2 measuring a distance between the shot motion activity descriptor and a set of
3 template motion activity descriptors.

- 1 10. The method of claim 1 wherein the distance is a semi-Hausdorff distance.
- 1 11. The method of claim 1 wherein the template motion activity is modeled by a
2 discrete function.
- 1 12. The method of claim 1 wherein the template motion activity is modeled by a
2 continuous function.
- 1 13. The method of claim 12 wherein the continuous function is a mixture of
2 Gaussian distributions.
- 1 14. The method of claim 1 further comprising:
2 extracting a plurality of training motion activity descriptors from sampled
3 frames of a training video including a plurality of training shots, each training shot
4 including a training talking head; and
5 combining the plurality of training motion activity descriptors into the
6 template motion activity descriptor.
- 1 15. The method of claim 1 further comprising:
2 segmenting the video into the plurality of shots using the motion activity
3 descriptors.
- 1 16. The method of claim 1 further comprising:
2 retaining only talking head shots.